# Competence and contributions to performance in belief-desire reasoning: Further evidence from the 'look first' task across two cultures

Amir Amin Yazdi [1], Tim P. German [2], Margaret Anne Defeyter [3] and Michael Siegal [4*]

1. Department of Psychology, Ferdowsi University, Mashad, Iran
2. Department of Psychology, University of California, Santa Barbara, U.S.A.
3. Division of Psychology, University of Northumbria, U.K.
4. Department of Psychology, University of Sheffield, U.K.

Acknowledgments: This paper was prepared with the support of the Iranian Ministry of Education and a grant from the Leche Trust (AAY), as well as support from the UCSB Academic Senate (TG), and a grant from the University of Northumbria (MD). Special thanks are due to Netalie Levi whose undergraduate project assisted in developing the studies reported here, to Veronica Guzzardi for drawing the pictures for Experiment 3, to Louisa Chan Egan and Danielle Truxaw for pilot testing of the picture task, and to Rosemary Firth, Jeanet Ingwersen, and Richard Parker for collecting the data reported in Experiment 3.

* Correspondence regarding this paper should be addressed to Michael Siegal, Department of Psychology, University of Sheffield, Western Bank, Sheffield S10 2TP, UK; e-mail: M.Siegal@Sheffield.ac.uk

**Abstract**

A recent meta-analysis of false belief task performance (Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory mind development: The truth about false-belief. *Child Development*, 72, 655–684) confirmed that there is change in false belief task performance across the 3-5 year age range. The analysis identified several 'performance' factors that influence task success, including manipulations that highlight the salience of the initial belief content (e.g. by asking where Sally will look *first* for the marble). Because a proportion of variance in performance remained unexplained even when identified performance factors were controlled for, the authors concluded that children's improvement is the result of conceptual change. Further, the analysis showed that manipulations such as look first improve performance only in children who are in the older part of the 3-5 year range, and thus plausibly operating with a 'transitional' theory of mind. Here, we present three studies systematically investigating the 'look first' manipulation which showed that: i) the advantage for the look first question can be demonstrated in children across different cultures, ii) look first has an effect that is additive to the improvement with age; there is no interaction such that older children gain more benefit from younger children, iii) performance in younger children can be, but is not always, elevated to levels that are statistically above 'chance. These results are discussed in terms of models of belief-desire reasoning in which *both* conceptual competence *and* performance factors play central roles.

## 1. Introduction

A major achievement in the domain of mental state reasoning concerns the capacity to solve problems in which the action of a character with a *false* belief must be predicted. For example, in the 'Sally-Anne' task (Baron-Cohen, Leslie, & Frith, 1985), children are told about a story character (Sally) who has a false belief about the location of a marble. The character is described as having placed the marble in a box but, when she is away, another story character (Anne) moves it into a different location. The test question concerns where Sally will look for the marble. Research has typically shown that 4-year-olds often succeed on such "standard" false belief tasks, whereas 3-year-olds do not (see Wellman, Cross, & Watson, 2001; for a review). The underlying mechanism that permits a shift from incorrect to correct responses on false belief tasks between these ages is still controversial (Bloom & German, 2000; Moses, 2001; Scholl & Leslie, 2001).

Two main accounts aim to explain the improvement across this age range. According to a meta-analysis carried out from a "theory-theory" perspective by Wellman et al. (2001), performance on false belief tasks increases from a level below what would be expected by chance at 3 years to an above chance level at 5 years, pointing to a conceptual change in children's understanding of mind. From the 'theory-theory' viewpoint, 3-year-olds' failure on false belief tasks can be been attributed to a conceptual deficit in their 'theory of mind' (ToM, Perner, 1991, 1995; Wellman, 1990; Wellman et al., 2001). In the light of research showing that most children at age of 4 years are able to answer correctly on false belief tasks, it is maintained that the onset of ToM reasoning is not generally achieved sooner than the age of 4 years and that there is a substantial development in ToM during the preschool years that permits children to understand the nature of false beliefs via the construction of a theoretical understanding of what beliefs are (Gopnik, 1993, 1996; Gopnik & Astington, 1988; Gopnik & Wellman, 1992, 1994; Perner, 1991, 1995; Wellman, 1990; 2002; Wellman et al., 2001).

However, the theory-theory account has been challenged by theorists suggesting that mental state reasoning is underwritten by a modular mechanism that is domain specific and reliably developing, and which forms the basis of the capacity to first attend to and then learn about mental states (German & Leslie, 2000, 2001; Leslie, 1994a, 2000; Scholl & Leslie, 1999, 2001). On this account, theory of mind concepts do not embody *theoretical* understandings of mental states, but rather serve to token instances of the relevant mental state in the world; they allow children to attend to the properties of believing, desiring and pretending, and thus learn about them (see German & Leslie, 2000, 2001; Leslie, 2000, Leslie, German, & Pollizi, in press, for extended discussion; Fodor, 1998, for a more general treatment of concepts).

On this latter story, if mental state concepts are early and reliably expressed, coming on line in typically developing children, at the latest, sometime within the second year (Leslie, 1994b, 2000) as indexed by the advent of pretend play (Leslie, 1987, 1994b), then something else must account for the shift in performance in the ability to pass false belief tasks at about age four. Under the 'early competence' view, the account that is offered for this developmental pattern proposes that gradual increases in processing capacities allow children to overcome features of belief desire reasoning problems that cause consistent errors (e.g., the tendency to attribute a true belief in the false belief task; Bloom & German, 2000).

There are increasing numbers of findings demonstrating that false belief performance in young children is affected by manipulations of such task demands (Birch & Bloom, 2003; Freeman & Lacohée, 1995; Happé & Loth, 2002; Mitchell & Lacohée, 1991; Roth & Leslie, 1991; Siegal & Beattie, 1991; Surian & Leslie, 1999; Zaitchik, 1991), as well as recent demonstrations that increases in the potency of the same kinds of

factors can cause problems for older children capable of solving 'standard' false belief tasks (Cassidy, 1998; Friedman & Leslie, in press; Leslie & Pollizi, 1998; Leslie et al., in press; Niehaus, Barrett & German, in preparation) and even affect adults' capacities to solve such problems (Birch & Bloom, in press; German & Hopkins, in preparation). Moreover, there are also large numbers of findings linking false belief performance with a number of more general 'executive' abilities, including inhibitory processing (Carlson, Moses, & Breton, 2002; Carlson, Moses & Hix, 1998; Carlson & Moses, 2001). These findings have engendered different specific performance accounts, but all share the basic notion that there will be manipulations of the standard false belief task that will have profound affects on performance, suggesting that conceptual change may play little role in explaining this particular developmental shift in the domain of theory of mind.[1]

Note that 'theory theory' accounts of mental state reasoning and the development thereof also acknowledge a role for performance demands. However, under the conceptual change view such factors play only a peripheral role. As Wellman et al. put it:

> Performance on a cognitive task reflects at least two factors: conceptual understanding required to solve the problem ("competence") *and other non-focal cognitive skills* (remembering key information, focus attention, comprehend, answer questions) required to access and express understanding (performance)." (Wellman, et al., 2001, p656-657, italics added).

The meta-analysis conducted by Wellman et al. (2001) documents the importance of performance factors in the solution of false belief tasks. However, because there was no single performance factor that 'removed' the difficulty for the younger child, Wellman et

[1] A full and extended discussion of the various performance accounts is beyond the scope of the current paper. The interested reader is referred to the excellent discussion in Moses & Carlson (2001; Carlson et al., 2002) introducing the distinction between 'expression' and 'emergence' accounts of the relationship between executive function accounts and 'theory of mind'. For a different take on this same debate, see

al., concluded that the 'truth' about false belief is that conceptual change must play a role in explaining the shift between ages three and four. Moreover, Wellman also claims that some putative performance manipulations are effective only for children at the older end of the 3-4 year range, and thus already close to achieving the conceptual change – i.e. who are in the 'transitional' stages of re-organizing their conceptual system in the appropriate way (Gopnik, 1996; Gopnik & Wellman, 1992, 1994). In the current investigation, we address these two conclusions by focusing on the role of one particular performance manipulation on children across the entire 3-4 year range.

### 1.1 The 'look first' manipulation

One of the first demonstrations of a performance manipulation with a profound effect on children's success in solving false belief problems was reported by Siegal and Beattie (1991). They designed two experiments to investigate responses on false belief tasks in terms of children's awareness of the purpose and relevance of test questions about the beliefs of a story character. They hypothesized that when an experimenter in the experimental setting asks questions such as "When will Sally look for the object?" 3-year-olds may not share the experimenter's intention that the question refers to how a person with a false belief will initially be misled. Instead, children may assume that the question refers to where Sally must or should look to find the object.

For this reason, Siegal and Beattie (1991) changed the question in the false-belief task from "Where will (the character with the false belief) look for the object?" to "Where will the character look *first* to find the object?" They argued that, as a consequence, young children are more likely share the purpose and relevance of experimenter's question. Their results showed that by adding the qualifier "first", 3-year-olds' performance improved significantly to a level that was above chance.

---

Perner & Lang, 2000). For a discussion of the relationship among difference performance accounts, see Leslie et al., in press).

The 'look first' advantage has been replicated, and also recently characterized as achieving its facilitating role in terms of introducing a reduction in the executive inhibitory demands present in the standard version of the task (Surian & Leslie, 1999; Leslie et al., in press). In what follows, we adopt the theoretical framework of Leslie & colleagues (German & Leslie, 2000, 2001; Leslie, 1994a, 2000; Leslie & Thaiss, 1992; Leslie, German & Pollizi, in press). Under this view, the early developing competence is supplied by a modular mechanism (the **T**heory **o**f **M**ind **M**echanism, or ToMM) that supplies basic mental state concepts (e.g. *belief*, *desire*, *pretend*) that allow the child to attend to behavior and infer the underlying mental state that gave rise to it. ToMM supplies a small number of candidate belief state contents for a given stream of behavior, and always includes a true belief content as a default. The true belief content is always supplied because in the absence of any other information, the only constraint on the protagonist's belief is the current true state of affairs (from the point of view of the attributer).

In most cases, the default attribution will not pose a problem because most of the time people's beliefs *are* true. From the perspective of conversational understanding between speakers and listeners, speakers normally abide by maxims or rules such as "Tell the truth and avoid falsehood" that prompt listeners to assume the truth of beliefs (Grice, 1975). But in some instances, speakers depart from such rules and so induce false beliefs in others. In such unusual situations where a false belief has been acquired, this default attribution must be *inhibited* and an alternative non-factual content for the belief selected instead. These performance demands are assumed to be handled by another mechanism, the Selection Processor (SP), which is a kind of executive function more domain general in character (at least, penetrable by more information than is ToMM), which may not necessarily be specific to the domain of belief desire reasoning (see e.g. Leslie & Thaiss, 1992; Roth & Leslie, 1998), though it is also *possible* that the inhibition and selection process is to some extent domain specific (see Leslie et al., in press).

Using this analysis, Surian and Leslie (1999) explain how the "look first" cognitively improves children's performance on false belief tasks. They argue that, in belief attribution, belief contents that are true are attributed *by default*. In a false belief task, the default true-contents leads to a wrong answer and therefore needs to be *inhibited*. By increasing the salience of the first location as the possible content of Sally's belief, the "look first" question reduces the need for inhibition. Therefore the "look first" false belief task places less of a load on the SP and enhances children's performance in comparison with their performance on standard false belief tasks. In this analysis, children as young as 3 years are able to represent beliefs. Their difficulty on standard false belief task results from limitations in access to attentional-conversational resources that gradually develop over the preschool period and beyond.

**1.2 Look first and meta-analysis**

Wellman et al.'s meta-analysis deals only with findings on look first (and similar manipulations aimed at increasing salience of the false belief content in other false belief tasks) published up to January 1998 (e.g., Freeman & Lacohée, 1995; Lewis & Osborne, 1990; Siegal & Beattie, 1991). The analysis excludes more recent results showing that explicit questioning on false belief tasks does enhance 3-year-olds' performance (Joseph, 1998; Nelson et al., 2003; Surian & Leslie, 1999).

Wellman et al., as noted above, maintain that "temporal marking" manipulations, as they term them, do not change the basic developmental pattern of false belief responses in preschoolers. In terms of the theory-theory account, performance manipulations only facilitate ToM performance once children at the older end of the 3- to 4-year-old spectrum are in the pre-conceptual change transition period – and thus have essentially the conceptual resources to answer false belief tasks correctly (this has been articulated as them having developed belief as an 'auxiliary' hypothesis; Gopnik & Wellman, 1994). Therefore, only among older 3-year-olds and young 4-year-olds, should

"look first" questions significantly increase correct responses on false belief task to a level above what would be expected by chance. According to this analysis, there should be an age × questioning format interaction effect in children's responses to standard and "look first" condition of false belief tasks (Figure 1).

---

Insert Figure 1 about here

---

In contrast to the theory-theory position, the hypothesized ToMM/SP pattern of children's responses on standard and "look first" false belief task suggests that there should be a relatively stable facilitating effect on false belief performance across the 3-4 year range. Scholl and Leslie (2001) outline the logic in more detail, but basically the look first manipulation can be thought of as adding a 'constant' amount of help to the inhibitory problem, which should increase the probability of success on a given trial by some measurable amount. This is depicted in Figure 2, showing main effects for age and condition (standard versus "look first") and no interaction. According to this account, reduction of performance demands through look first questioning leads to increased success on false belief tasks at all ages across the relevant age range.

---

Insert Figure 2 about here

---

In the studies we report here, we sought to adjudicate between these alternatives using first a large sample of children that would permit a more detailed analysis of response patterns. In the follow up studies we addressed two further questions i) whether the advantage for look first questions over standard questions was the result of false positives – success achieved via general biases toward responding to the first location, irrespective of what was the character's belief, and ii) whether the effect of look first can be demonstrated using a within subjects experimental design, and specifically in a group

of children who are at the younger end of the 3-4 year old range (and thus not plausibly in the 'transitional' stage of conceptual development prior to theory change).

At the same time, since previous studies on look first effects have been limited to industrialized Western cultures, our aim was to determine whether look first effects emerge in children living in the culture of a developing country. In our first experiment, we aimed to re-examine the Western pattern of children's responses on standard and "look first" false belief tasks by attempting to replicate Siegal and Beattie's (1991) findings in Iran.

## 2. Experiment 1

### 2.1 Method

#### 2.1.1 *Participants*

These were 194 children were divided into two age groups of 90 3-year-olds (mean age= 3 year, 5 months, range= 36 to 47) and 104 4 year-olds (mean age = 4 year, 4 months, range= 49 to 60). The children attended preschools located in middle class areas of Mashad, Iran. All children had parental permission to participate. In addition, nine children, two 3-year-olds and seven 4-year-olds, answered incorrectly on reality control questions. The data from these children were discarded from final analyses.

#### 2.1.2 *Materials and procedures*

As in Siegal and Beattie's (1991) Experiment 1, in each age group, the children were randomly assigned to one of two conditions: standard or look first. Each child was tested individually on two stories. In order to make use of false belief tasks with Iranian children, the stories were initially translated into the Persian language and then the Persian version was translated back into English to ensure that the Persian version corresponded with the English one. For the children in the standard condition, the stories were similar to those used in standard false belief tasks (Wellman & Bartsch, 1988): (1)

"Maryam wants to find her kitten. Maryam thinks her kitten is in the kitchen. Maryam's kitten is really in the bathroom. Where will Maryam look for her kitten? Where is it really?" (2) "Ali wants to find his puppy. Ali's puppy is really in the kitchen. Ali thinks his puppy is in the bathroom. Where will Ali look for his puppy? Where is it really?" The locations of the pets in the stories were counterbalanced across subjects so that, for half the children, Maryam was said to believe that her kitten was in the kitchen (real location: bathroom) and Ali to believe that his puppy was in the bathroom (real location: kitchen); for the other half, the real and believed locations were reversed. The procedure for the look first condition was identical to that of the standard condition except that the children were asked, " Where will Maryam (Ali) look first for her (his) kitten (puppy)?"

Within each age group, the order of the stories was counterbalanced, as were the orders of the premises concerning beliefs and reality. During the story presentation, the experimenter referred to miniature figures of a girl, boy, kitten, and puppy, and to locations inside a large wooden doll house (60 X 60 cms).

**2.2 Results**

Children's responses on false belief tasks were scored on a 0-2 scale with each correct answer receiving one point. The scores were analyzed in a 2 (Age) X 2 (Condition) factorial ANOVA. There were significant main effects for age, $F$ (1, 194)= 6.52, $p < .01$) and condition, $F$ (1, 194)= 6.05, $p < .01$. The age X condition interaction effect was not significant, $F$ (1, 194)= .65, $p = .41$. The 4-year-olds outperformed 3-year-olds with mean of 1.31 out of maximum possible score of 2 (65% correct) compared to 3-year-olds' mean score of .65 (32% correct). Children questioned in the look first condition outperformed in the standard condition. The mean score in the look first condition was 1.32 out of a maximum possible of 2 correct (66% correct) compared to the mean in the standard condition of .68 (34% correct).

Consistency in the children's responses on false belief tasks is shown in Table 1. Children in the standard condition were consistent in answering incorrectly to false belief tasks. Of the 45 3-year-olds in the standard condition, 4 were correct on both stories, 5 on one story, and 36 on neither, a performance level that is far below chance ($p$ <01, binomial test). By contrast, those in the look first condition performed at an above chance level. Out of 45 3-year-olds, 17 were correct on both stories, 12 were correct on one story, and 16 on neither. Assuming a 25% chance success rate, the numbers of children answering correctly on both stories exceeded chance level expectations, z = 1.807, p = .0307. The 4-year-olds in both conditions were consistent in answering correctly. In the look first condition, out of 53 4-year-olds, 40 were correct in both stories, 4 were correct on one story, and 9 on neither. Comparable numbers for the standard condition were 22, 9, and 20.

---

Insert Table 1 about here

---

Altogether the children's responses conformed to the pattern predicted by the ToMM/SP position (Figure 3). Even at the youngest ages, there was a significant effect for condition (Figure 4). Of the 22 children aged 36 to 41 months in the standard condition, 20 children answered incorrectly on both tasks; one child was correct on both tasks and another was correct on one of the two tasks. By contrast, of the 25 children aged 36 to 41 months in the look first condition, only 11 children answered incorrectly on both tasks; 6 children were correct on both tasks and 8 were correct on one task.

---

Insert Figure 3 about here

---

---

Insert Figure 4 about here

---

**2.3 Discussion**

Consistent with the ToMM/SP hypothesis, the results support the notion that the "look first" improves both 3- and 4-year-olds' performance on false belief tasks. This is in contrast with the theory-theory account in which 3-year-olds' performance is supposed to not be improved in the "look first" condition.

Concerning the role of "look first" on children's performance on false belief tasks, the findings are consistent with the results obtained from Siegal and Beattie's (1991) Experiment 1 in which children gave 71% correct answers in the look first condition and only 30% correct answers in the standard condition. The comparable results for the current experiment were 57% correct answers in the look first and 32% in the standard condition. In both Siegal and Beattie's Experiment 1 and the current experiment, children performed better in the "look first" than the standard false belief tasks. In both Siegal and Beattie's Experiment 1 and the current experiment, the 3-year-olds' performance on standard false belief tasks was far below what is expected by chance. While 3-year-olds in the look first condition who were in Siegal and Beattie's (1991) Experiment 1 performed at an above chance level, the performance of 3-year-olds in the current experiment was no greater than chance.

Unlike the results of Siegal and Beattie (1991), the results of the current experiment also indicate the main effect of age demonstrating that 4-year-olds outperformed 3-year-olds' performance on false belief tasks. Iranian 3-year-olds' performance on both the standard and look first conditions in this experiment was slightly worse than that of the Western children tested by Siegal & Beattie. The difference may be due to the fact that the Iranian children (mean age = 3 year, 3 months) were on average younger than Western children (age mean = 3 year, 8 months) in Siegal & Beattie's (1991) Experiment 1.

However, it might be argued that the experimenter's use of the term "*first"* might have somehow have signaled to children that they should give answers that are contrary to the real location of the desired object. If so, the children then could succeed on the task without attending to the character's beliefs. The purpose of the next experiment is to replicate the results of Siegal and Beattie's (1991) Experiment 2 in which they used a true belief control task in order to guard against the possibility of false positive responses. On the one hand, if children interpret the look first question as generically implying that they should give an answer different from the real location of the object, then they should respond correctly in the look first condition of false belief tasks but not in the look first condition of true belief tasks. On the other hand, if they give correct answers on both look first false and true belief tasks as in Siegal and Beattie's (1991) Experiment 2, they would be interpreting the term "first" in a manner sensitive to the character's belief state, ruling out the false positive interpretation. By aiming to replicate the results of Siegal and Beattie's (1991) Experiment 2 (see also Surian & Leslie, 1999), we sought to investigate whether "first" creates its effect by inducing false positive responses or whether it truly leads to better belief calculation.

## 3. Experiment 2

### 3.1 Method

#### 3.1.1 *Participants*

The subjects were 22 3-year-olds with a mean age of 3 year, 6 months, range = 3:1 to 3:11. An additional 11 children, 7 in the "look first" condition and 4 in the standard condition, answered reality control questions incorrectly and were discarded from final analyses. The children attended preschools located in middle class areas of Mashad, Iran. All had parental permission to participate. No child had participated in Experiment 1

### 3.1.2 *Material and procedures*

As in Experiment 1, the children were randomly assigned to one of two groups: the standard condition and the look first condition. For the standard condition group (n = 13), the tasks were to predict the actions of a story character based on a true belief (TB) or a false belief (FB). For the look first condition group (n =9) the tasks were the same but involved predicting where the story character would look first either in regard to a TB or FB.

The TB story was "Maryam wants to find her kitten. The kitten lives in two rooms: the garage and the lounge. Maryam thinks her kitten is in the garage and now it really is in the garage. Where will Maryam look for her kitten? (test question) Where is the other room that the kitten lives in?"(control question). The false belief story was similar to that used in Experiment 1: "Ali wants to find his puppy. Ali's puppy is really in the kitchen. Ali thinks his puppy is in the bathroom. Where will Ali look for his puppy? Where is it really?" The content of the TB and FB stories in the standard condition was identical to TB and FB stories in the look first condition respectively except that the children in the "look first" condition were asked, "Where will Maryam (Ali) look first for her kitten (his puppy)?"

The locations of the pets in the stories were counterbalanced across subjects so that, for half the children, Maryam was said to believe that her kitten was in the lounge (real location: lounge) and Ali to believe that his puppy was in the kitchen (real location: bathroom); for the other half, the real and believed locations were reversed. The orders of the premises within the stories were also counterbalanced. The same materials as in Experiment 1 accompanied the stories.

## 3.2 Results

Table 2 shows children's responses to test questions in the standard and look first conditions. Children in the look first condition outperformed children in the standard one.

While 7 of the 9 children in the "look first" condition recognized that the character would look in the wrong location when the belief was false, the comparable numbers in the standard condition was only 2 of 13. The difference between the groups was significant ($p < .05$, Fisher exact probability test). By contrast, there was no significant difference between children's performance on standard and "look first" TB tasks. In this condition, most of children responded that the character that had a true belief would look where he or she thought the desired pet was located. Of the 13 and 9 children in the standard and look first conditions respectively, the numbers correct on the TB tasks were 13 and 7 respectively.

---

Insert Table 2 about here

---

The children's performance on FB and TB tasks in look first condition, however, was similar. Of the 9 children in the look first condition, the numbers correct on the TB and FB tasks respectively were 7 and 7. The similarity might be explained by the fact that children on the FB task performed at ceiling.

**3.3 Discussion**

The results support the hypothesis that children's better performance on the look first condition does not stem from inducing false positives – correct responses that are caused by a bias to select the first location irrespective of the belief state of the character. If a general 'first location bias' had been introduced by the look first procedure, it should be applied whether it is the right answer (as in the case of the false belief stories) or the wrong answer (as in the true belief control stories). The results of Experiment 2 replicate the results of both Siegal and Beattie's (1991) Experiment 1 and Surian and Leslie (1999), as well as those of Experiment 1 here in showing that children often respond correctly when asked to predict the initial behavior of a story character with a false belief.

These results do not support the alternative hypothesis that children misinterpret the word *first* and perceive the question " Where will he look for first?" to require an answer that is other than the real location of the desired object. If a general bias to respond toward the empty location were being introduced, children ought to have made errors in the true belief control condition with the look first question – a result that did not occur in this experiment.

The present Experiments 1 and 2 show that children appear to be helped in attending to the first location as a possible target content for the character's belief. This result was initially characterized as stemming from a 'clarification' in the experimenter's intended question; the look first version of the task lowers the likelihood that the 3-year-old will interpret the test question not as "Where will the character look (first) for the object?" as intended by the experimenter but instead will gloss it as something like: "Where should the character *to find the object*?" (Siegal & Beattie, 1991). Within the ToMM-SP framework, we can further articulate the mechanism via which this clarification might work, namely, in increasing the salience of the false belief content such that lower inhibitory demands are made during the selection process (Leslie et al., in press; Surian & Leslie, 1999).

In Experiment 3, we attempted to provide further evidence for the effect of the look first manipulation, and its role across the 3-4 year old age range. In this experiment , we were particularly concerned with showing that the effect of look first is stable and appreciable even among the youngest 3-year-old children. Thus, we tested a sample of young three year olds only, and compared them to a group of young four year olds. This experiment thus provides a focused test of Wellman et al.'s claim that the look first manipulation works only for children who are in the transitional period immediately prior to conceptual change. Secondly, we tested the effect of the look first manipulation with a couple of changes to the materials and procedure of the false belief task, and in a group of children from a western cultural setting.

The first change was to follow Surian and Leslie (1999) and test the look first procedure in a task modeled on the standard change of location task (Baron-Cohen et al., 1985), rather than with tasks modified from Wellman & Bartsch (1988). We chose a picture book version of the Sally-Anne task for this purpose.

The second change to the procedure in this experiment was that we dealt with the issue of false positives by using a true belief task (as reported in Experiment 2). However, we used this task as a screening task, such that only children who solved both 'think' and 'look first' versions of the true belief task were admitted to the false belief study. In this way, we were able to be sure that none of our subjects was susceptible to general biasing effects of the 'look first' question.

The final change was to assess the effect of look first within the same subjects. Because of possible interference caused by asking both 'look' and 'look first' questions to the same children in quick succession, we achieved the within subjects procedure by using a 'think' question as our standard assessment of false belief, and comparing this to 'look first'.[2] We speculated that in a within subjects design, any advantage for look first might have some carry over effect for the think question, if children are induced by look first into being more likely to calculate a false belief content, and on the assumption that this information is available to a subsequent belief calculation process (one might conceive of this as a kind of 'priming'). Similarly, receiving the think question first (and calculating it incorrectly) might have a lingering effect on the activation of the true belief content, making it more highly activated, and reducing the effect of look first manipulation on easing the subsequent action prediction.

[2] While there are no appreciable differences between the 'think' and 'look' versions of the false belief task (Wellman et al., 2001), it is important to note that this holds only for false belief tasks that also involve a desire to *approach* the object. A recent discovery is that action prediction is considerably harder than 'think' when the character's desire is to *avoid* the object in question (Cassidy, 1998; Leslie & Pollizi, 1998; Leslie et al., in press).

## 4. Experiment 3

### 4.1 Method

#### 4.1.1 Participants

The participants were 70 children aged between 3 years 0 months and 4 years 3 months. Of these children, eighteen failed to solve the TB screening task, either by failing the look first question (3 children), the think question (11 children) or both (4 children), resulting in a final sample of 52, divided into a 'young three' year old group (N = 22, mean age 3-3, range 3-0 to 3-5) and a 'young four' year old group (N = 30, mean age 4-3, range 3-11 to 4-7). There were approximately equal numbers of boys and girls in each age group, and the children's ethnicities reflected a typical urban northern city in the UK. All children had English as a first language.

#### 4.1.2. Materials and procedure

Two picture book tasks were produced, each with 7 pages, conveying the main events of each task (true belief and false belief). The pictures, stories, questions and questions used in Experiment 3 are presented in Figure 5. Children were randomly assigned to receive the either the think question first or the look first question first for both the true belief screening and false belief task.

Children were tested individually, seated at a low table in a quiet area of their preschool. The children were first presented with the true belief screening task. The story was presented one picture at a time, with the text (see Figure 5) accompanying each story read out at that time. If children failed a control question (memory or reality), they were corrected, recycled through the story and asked again. If they failed a second time, this would have resulted in them being eliminated from the story, but in fact no child failed a second time. Children who solved the true belief task (defined as correctly answering both think and look first questions correctly)were included in the main study. After a

short break in which the children were allowed to return to play in the class, those eligible for the false belief task were recalled and given the false belief story.

---

Figure 5 about here

---

**4.2 Results**

Children were scored correct if they answered that Sally would look for her marble under the bed (look first question) or thought that the marble was under the bed (think question). The number of children (and percentage of children passing the look first and think versions of each task) is shown in Table 3. There was a significant advantage of age for performance on the think question ($X^2_{(1)} = 9.93$, $p < .002$, one tailed), but this difference was not significant for the look first question ($X^2_{(1)} = 1.98$, N.S.).

---

Table 3 about here

---

The relation between performance on the think question and performance on the look first question is shown in Table 4. This shows that there were more children who passed the look first question while failing the think question than who showed the opposite pattern, a pattern that was significant for both 3-year-olds (McNemar binomial, N = 10, k = 0, $p = .002$, one tailed) and for 4-year-olds (McNemar binomial, N = 5, k = 0, $p = .036$, one-tailed).

Finally, we investigated whether the advantage for look first was in any way affected by the order in which the questions came. This analysis revealed that while there was a tendency for children to solve the think and look first questions slightly better

when the look first question was asked first (think question 26% correct when asked first, 48% when asked second; look first question 72% when asked first, 57% when asked second), this difference was only marginally significant for think question, by one tailed criterion ($X^2$ (1) = 2.67, $p$ = .056) and not significant for the look first question ($X^2$ (1) = 1.41,.).

---

Table 4 about here

---

**4.3 Discussion**

The results of this experiment were clear. There was an advantage for 'look first' over 'think' versions of the false belief task that was demonstrated i) within the same subjects, and ii) within a group of subjects at the lower end of the 3-4 year age range. None of the three-year -old children in this experiment were drawn from an age range close to the hypothesized 'transitional' periods prior to conceptual change (Gopnik & Wellman, 1992, 1994). Nevertheless, there was an advantage for look first over think in this youngest group that was highly significant, while the advantage for the young four-year-old group, who overall were far from ceiling on the standard 'think' question (57%), was more modest in the look first condition (improved to 73%). Though this was a significant improvement via a one tailed test, it was not as dramatic as that observed in the younger group, who improved from 14% performance on the think question to 55% on look first.

Thus, though effects of look first were demonstrated among both three and four year olds in Experiment 3, this experiment comes the closest to showing an interaction between age and the effect of this performance factor. However, it is not the interaction that is predicted on Wellman et al.'s (2001) argument about the effect of this task

manipulation, but rather hints at an interaction in which the youngest children gain more than their older peers

Finally, we note with interest a slight, though statistically weak order effect in this within-subjects experiment, suggesting that the advantage for look first tended to facilitate children's performance on a subsequent think question. Receiving a think question first showed a smaller trend toward hindering the subsequent look first question. While the pattern of results is consistent with our predictions concerning the lingering effect of information in the processing system on subsequent question performance, it reached only marginal significance in the current dataset. Therefore we refrain from speculation pending replication.

**5. General Discussion**

The results of the present research support the pattern of performance on "look first" questions hypothesized by the ToMM/SP theory, as depicted in Figure 2. Recall that according to the ToMM/SP account, 3-year-olds' difficulty in standard false belief tasks results from a performance limitation in overcoming a bias to attribute the default true belief content in false belief situations. If the failure on false belief problems between age 3 and 4 is explained this way, then the need to propose machinery that realizes a conceptual shift at this age goes away, and a major obstacle to the idea that belief-desire reasoning is underwritten by a modular reliably developing mechanism is removed.

Under this view, the early developing mechanism (ToMM) comes online during the second year of life and supplies basic attitude concepts (e.g. belief, desire, pretend) that allow the child to attend to behavior and interpret it as emanating from mental states. However, because ToMM is limited in its ability to access information from all parts of the cognitive system, it relies on the functioning of a further executive selection process (SP) to select among the range of candidate mental state contents calculated on the basis of the perceptual description of the behavior. The selection processor increases across

development in its efficiency at the inhibition-selection process, producing a change in performance on standard tests of false belief. Versions of the false belief task that introduce reduced selection demands (such as look first) create situations where children are probabilistically more likely to generate the correct response (Siegal & Beattie, 1991; Surian & Leslie, 1999). Versions of the false belief task that introduce additional executive or inhibitory demands (such as combining the false belief with a desire to avoid the object) create situations where children are probabilistically less likely to generate the correct response (Cassidy, 1998; Leslie & Pollizi, 1998; Friedman & Leslie, in press; Leslie et al., in press; Niehaus, Barrett, & German, in preparation). Interestingly, one recent demonstration shows that the facilitating effect of 'look first' on the performance of 3-year-olds on standard false belief problems, also extends to reducing problems for 4-year-olds on the false belief tasks that have increased inhibitory demands (Leslie et al., in press).

The current studies provide more evidence on the specific nature and profile of the effect that 'look first' has on performance on false belief reasoning across the 3-4 year period, and in a culture as yet unstudied by researchers interested in belief-desire reasoning. Specifically, it shows that the effect of look first is *not* restricted to those children who are 'closest' to achieving a conceptual change (e.g. only older 3 year olds or young four year olds). Across all three studies, there was an appreciable advantage of look first over the standard form of question that extended to the youngest children in the sample. Moreover, in one study (Experiment 2) the advantage for younger children exceeded 'chance' (see below), while in another the advantage was, if anything, better in young 3-year-olds than it was in young 4-year-olds (Experiment 3).

This specific pattern of results is not consistent with the conceptual change explanation of 'look first' and similar manipulations offered by Wellman et al., (2001) in discussing their meta-analysis. The pattern that Wellman et al show in that analysis (where look first and similar manipulations helps only older children) is likely skewed by

the fact the 1998 cut off date excluded several 'look first' type manipulations (Joseph, 1998; Nelson et al., 2003; Surian & Leslie, 1999; see Scholl & Leslie, 2001).

Our results address several issues that are prominent in debate surrounding competence and performance in belief desire reasoning: first, the notion that improved performance on false belief is meaningful only if performance is elevated above 'chance', and second, how the 'look first' manipulation illuminates the question of how apparently different performance factors can be captured in a common framework.

**5.1 Competence, performance and 'chance'**

Considerable controversy in the 'theory of mind' literature has surrounded the proposal that demonstrating that a task manipulation leads to some improvement in false belief performance is itself not important unless the improvement in performance is to a point that is above 'chance' performance (Wellman et al., 2001), where chance is assumed to be 50%, on the basis of the fact that there are typically two responses in the false belief task.

While we agree that demonstrating that performance statistically exceeds that expected if children were 'guessing' suggests that the improved performance is not attributable to the manipulation causing children to respond at random, it is less clear that performance that does not exceed chance is always best explained by assuming that children are 'just guessing'. On the first point, it is important to note that while performance that exceeds 'chance' suggests that children are not responding randomly *with respect to the response options available*, it does not in itself show that the performance has improved for non-trivial reasons. For example, if a set of results shows that alongside perfect responding on a false belief problem with a new manipulation, the same children *all fail* a true belief problem with the same manipulation, then this pattern of results suggests that the improvement is likely caused by some non interesting bias to

respond to one location. This is true, despite the fact that the performance in both conditions is significantly different from 50-50 'chance'.

With respect to the second point, we note that across the current studies children's performance was not elevated to 'above chance' levels in Experiments 1 and 3 (though it was in Experiment 2). However, it seems that there are a number of good reasons to reject the idea that any responding that does not exceed 50% (where there are two options) is best explained by assuming that the children were guessing.

First, and perhaps trivially, although it is rarely pointed out, there are no specified developmental theories of 'guessing', or when and how guessing is to be expected. For sure, there are numerous discussions of various biases that might affect children's performance on various tasks (such as a 'yes' bias to respond in the affirmative, Fitzley & Lee, 2003). The advantage of theories that posit that biases or response strategies, is that they generate hypotheses that can be tested. Over the extent of the literature, they can be subjected to the same rigorous test as other theories of cognitive development (such as conceptual change).

However, we know of no theoretical formulations that propose when and how children actually generate random responding. The notion that sometimes children simply guess, because its details are not articulated or specified, becomes simply a catch-all theory that has no support from independent evidence. To us, there does not seem any good reason to prefer such a theory to an articulated, testable one.

The second general consideration is that the idea that performance must be above chance to be meaningful is perhaps more usefully applied in cases where there is no background 'bias' toward one or other response. It is notable in the false belief task that children rarely respond randomly in standard versions of the task – indeed the fuss has all mostly been because children make such systematic errors. So if 50-50 responding is rarely seen in such tasks, on what basis is it a benchmark for determining that a novel result is important? Part of the reason may reflect the tendency within studies of

cognition in young children to frame successes and failures in terms of the child 'having' or 'not having' understanding of a concept. The notion that about half the children at a given age might get an answer right does not translate easily into this framework. One is tempted to worry whether they 'get it' or don't.

These problems do not arise if one considers 'the child' and 'understanding' as divisible entities. A manipulation might have some systematic effect on the functioning of one of the elements that contributes to performance, leading to a systematic shift in the likelihood of a given response (Scholl & Leslie, 2001). Thus, moving children away from a significant error or pattern of performance by some non trivial and statistically significant margin need not require that the child (as a whole) needs now to be thought of as 'understanding' the concept, where previously she did not. Instead, the manipulation has simply had an effect on one aspect of the information-processing procedures that the cognitive system performs during the execution of the problem.

Turning to the current results, there are several reasons to suspect that the improvement in the youngest children in Experiments 1 and 3 was not induced by the manipulation creating false positives by inducing bias toward the correct location in response to the 'look first' question.

First, all the children in Experiment 3 passed the true belief screening task prior to being admitted into the experiment. This gives us *prima facie* evidence to suppose that they were not susceptible to being biased by the manipulation, nor confused by it. The true belief tasks are a powerful control against the notion that manipulations induce false positives because of bias and or confusion. This is because the notion of a manipulation biasing children toward one response requires that the bias or confusion hold across different versions of the task. If not, then the 'bias' account must somehow explain how the children somehow know *not* to be biased on some occasions. Demonstrations of 'early' false belief understanding via looking time measures have made extensive use of

the true belief control to show that low level response biases do not account for successes that are demonstrated in false belief detection tasks (e.g. Clements & Perner, 1994).

Secondly, the results fit within a broad and increasing literature of demonstrations of successful false belief reasoning in younger 3-year-old children, which we review briefly below. While isolated results demonstrating success on false belief tasks that do not demonstrate performance in excess of 'chance' might be conservatively attributed to guessing, as the number of demonstrations increases, it becomes increasingly plausible to suggest that such performance shifts reflect something real in terms of their effects on the child's cognitive system. The notion that only manipulations that create 'full understanding' (assessed as better than chance responding) in 'indivisible' children, in our view, itself becomes the less conservative position.

**5.2 Integrating 'conversational' and 'executive inhibitory' performance factors**

It has been suggested that conversational understanding is a resource for computing answers on false belief tasks that becomes more efficient during development. This process has been advanced as an explanation for 3-year-olds' difficulty in false belief tasks – lack of success is attributed to resource limitations on children's ability to acknowledge the experimenter's purpose in asking the test question in an experimental setting (Siegal, 1997, 1999). At this level of analysis, the look first question acts as a conversational aid, providing a computational resource that helps 3-year-olds to process the appropriate intent behind the question and calculate the correct response. This notion has great intuitive appeal, because the manipulation can be captured as one that 'clarifies' things for the younger child, whose conversational resources are less sophisticated than those of the older child. The older child, in possession of greater conversational resources, does not need the clarification in the standard version of the task.

However, the ultimate aim of cognitive science, with its commitment to the computational theory of mind (see e.g. Fodor, 1975; Pinker, 1997), is to advance explanations of constructs such as 'conversational resources' and 'clarification' in terms of the framework theory of the processing of information. It is therefore instructive to make an attempt to capture the effect of look first in terms of models of the processing of information during belief desire reasoning (see Leslie et al., in press). Leslie and colleagues have advanced the only models of belief-desire reasoning in which the processes leading to successful performance have been specified (Friedman & Leslie, in press; Leslie & Pollizi, 1998; Leslie et al., in press). The major currency in which task success and failure have been cashed out under these models is in terms of manipulations affecting the ability to control selection of responses via inhibition (see also German & Leslie, 2000, 2001; German & Nichols, 2003; Leslie, 1994a, 2000; Leslie & Thaiss, 1992; Roth & Leslie, 1998; Surian & Leslie, 1999). [3]

Under this account, and following Surian & Leslie (1999), we can offer a mechanism through which look first plays its clarification role. The idea is that the look first question draws attention to the first location (i.e. the target of the false belief content) and renders it more salient than it otherwise would be. The question format thereby tends to reduce the salience *differential* between true- (default) and false-contents. The reduced differential in turn requires less inhibition to reverse its direction, allowing children with lower inhibitory resources a greater chance of success (see Leslie et al., in press). Note that the inhibition reduction account of 'look first' enhances Siegal and Beattie's account in providing a specific mechanism for how clarification might

[3] We restrict this claim to models of tasks assessing belief-*desire* reasoning (e.g. action prediction false belief tasks such as the Sally-Ann task). Barreau & Morton (1999) offer a plausible memory based model for performance in a task requiring belief attribution only (to the self).

work (Surian & Leslie, 1999). Such a mechanism has greater explanatory scope because it not only explains how clarification can occur for the three-year-old, but also why four-year-olds do not need the same in order to pass standard false belief prediction.

Extending this account, we note that the same general explanation can be applied to explain why, even though 4-year-olds do not need help in the standard approach desire versions of false belief, they have difficulty in versions of the false belief problem with additional inhibitory demands, such as where the action resulting from a false belief coupled with a desire to *avoid* the object must be predicted (Cassidy, 1998; Friedman & Leslie, in press; Leslie & Pollizi, 1998; Leslie et al., in press). Note that without the inhibitory framework for conceptualizing performance factors, it is not obvious why 4-year-olds who have sufficient conversational resources to 'get the point' of the question for tasks with approach desire, should suddenly have problems for the very same question asked about a task with avoid desire.

Conceptualizing conversational resources in terms of the control of inhibition, however, makes sense of this otherwise puzzling finding, and as noted earlier, allows the prediction that 'look first' might provide relief for the four year old children who have difficulty with avoid desire, just as it provides relief for those 3-year-olds who otherwise would err in approach desire tasks. This is indeed the pattern of results that is seen (Leslie et al., in press), and moreover is a result that provides insight into determining which of the two different processing models of success proposed by Leslie & Pollizi (1998) is correct (see also Friedman & Leslie, in press).

The current paper is thus one of several recent attempts to integrate studies of the effects of various aspects of language development on the expression of mental state reasoning abilities. The interactions between language and 'theory of mind' comprise

effects of syntactic development (de Villers & de Villiers, 2000; de Villiers & Pyers, 2002), semantic development (Bloom, 2000; Happé & Loth, 2002;) and conversational pragmatic development (Woolfe, Want, & Siegal, 2002). While these approaches might been seen as importantly different from one another in their emphasis of different performance factors in explaining the pattern of mental state reasoning across the preschool period and beyond (see Leslie et al., in press, for discussion), there is important common ground that can be explored in attempting to capture them in a common framework, and thereby bring them together also with work demonstrating relationships with domain general measures of executive function (e.g. Carlson et al., 1998; 2002).

Such integrative attempts will have to capture several important findings that are beginning to emerge from the study of the relationship between language and mental state reasoning. In these investigations, impaired grammar does not appear to compromise belief-desire reasoning skills (Siegal, Varley, & Want, 2001); many young children are adept at semantics and syntax in terms of their understanding of the task structure of ToM measures but still do poorly (Custer, 1996; Woolfe et al., 2002). These findings must be reconciled with the effects of 'focused training' of certain kinds of linguistic skills (such as sentence complementation, DeVilliers & DeVilliers, 2000; DeVilliers & Pyers, 2002) on later belief desire reasoning performance (Hale & Tager-Flusberg, 2003; Lohmann and Tomasello; 2003). Although Hale and Tager –Flusberg (2003) and Lohmann and Tomasello (2003) have reported success at training ToM performance with exposure to instruction on sentence complementation, Ruffman et al. (2003) and Cheung et al. (in press) report evidence that ToM reasoning is related to general language ability rather than to specific aspects of syntax or semantics. Moreover, as Lohmann, Tomasello, and Meyer (in press) recognize, training studies on sentence

complementation may in fact involve exposure to discourse that may foster conversational understanding which in turn promotes success on false belief tasks.

### 5.3 Epilogue: The truth, the whole truth and nothing but the truth about false belief?

The publication of the meta-analysis of false belief task performance was under the subtitle of 'The truth about false belief" (Wellman et al, 2001). While the analysis is certainly informative, there is a danger in over extending the statistical conclusions that one is entitled to draw from such an enterprise (e.g. that there is abundant evidence for development in the performance at false belief across the 3-5 year period), into theoretical conclusions (e.g. that this development is best explained in terms of conceptual change). As pointed out by commentators on the analysis (Moses, 2001; Scholl & Leslie, 2001), while the extensive evidence presented in the meta analysis is *consistent* with conceptual change (e.g. there is variance in performance that is not explained by the known performance factors introduced into the 'best' model identified by Wellman et al.), this does not actually demonstrate conceptual change, nor is it inconsistent with 'early competence' accounts, as look first exerts a significant effect even on the false belief performance of young 3-year-olds in different cultures.

The current series of studies has tested one major specific prediction of conceptual change accounts for one of the performance factors discussed in the Wellman et al. (2001) meta-analysis – the effect of 'temporal marking' (Freeman & Lacohée, 1995; Lewis & Osborne, 1990; Siegal & Beattie, 1991), which included the look first task. This prediction, based on the output of the meta-analysis, was that only children toward the older age end of the 3-5 year range would benefit from temporal marking – children who might be thought of as in a 'transitional' period between one theoretical understanding of mind and the next (Gopnik, 1996; Gopnik & Wellman, 1992, 1994), or

who might be thought of as in something akin to a 'zone of proximal development' (Vygotsky, 1962) with respect to achieving the mature conception of belief.

The current series of studies demonstrates that even the statistical conclusion from the meta-analysis does not hold up; there is an advantage for look first over standard questions even among children in the youngest part of the 3-5 year range. This underlies one limitation of meta-analysis; one gets out only some version of what is put in, and if there are important datasets excluded, as will be inevitable with a cut-off date, then the meta-analysis will not reflect these datasets. The 'truth' about false belief, even in terms of a statistical description of performance (and irrespective of theoretical considerations) is thus at best an approximation that falls short of the whole truth.

We would also urge caution, following Scholl & Leslie (2001), in the interpretation of unexplained variance as providing support for a particular mechanism of development (e.g. conceptual change). Statistically, unexplained variance is just that – *unexplained*; it cannot *a priori* be claimed as support for one position over any other. It is our view performance factors should be treated not as 'artifacts' that should be 'removed' for best consideration of development in the domain of theory of mind, but rather as major features of information processing models of successful belief-desire reasoning. These factors are worthy of study and explanation in their own right, and in doing so, scholars may advance toward a truth about false belief that is believable.

**References**


Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37–46.

Barreau, S., & Morton, J. (1999). Pulling smarties out of a bag: A Headed Records analysis of children's recall of their own past beliefs. *Cognition*, *73*, 65–87.

Birch, S.A.J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science, 14*, 283-286.

Birch, S.A.J., & Bloom, P. (in press). The curse of knowledge in reasoning about false belief. *Cognition.*

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bloom, P., & German, T.P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*, B25–B31.

Carlson, M.S. & Moses, L.J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032-1053.

Carlson, S.M., Moses, L.J., & Hix, H.R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, *69*, 672–691.

Carlson, S.M., Moses, L.J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*, 73-92.

Cassidy, K.W. (1998). Three- and four-year-old children's ability to use desire-and belief- based reasoning. *Cognition*, *66*, B1–B11.

Cheung, H., Chen, H-C., Creed, N., Ng, L., Wang, S. P., & Mo, L. (in press). Relative roles of general and complementation language in theory-of-mind development: Evidence from Cantonese and English. *Child Development.*

Clements, W.A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, 9, 377–395.

Custer, W.L. (1996). A comparison of young children's understanding of contradictory representations in pretense, memory, and belief. *Child Development, 67,* 678-688.

DeVilliers, J.G., & DeVilliers, P.A. (2000). Linguistic determinism and the understanding of false beliefs. In (Eds) P. Mitchell and K.J. Riggs, *Children's reasoning and the mind.* (pp. 191–228). Hove, UK: Psychology Press Ltd.

DeVillers, J.G. & Pyers, J.E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false belief understanding. *Cognitive Development, 17,* 1037-1060.

Freeman, N. & Lacohée, H. (1995). Making explicit 3-year-olds' implicit competence with their own false beliefs. *Cognition, 56*, 31-60.

Friedman, O., & Leslie, A.M. (in press). Mechanisms of belief-desire reasoning: Inhibition and bias. *Psychological Science*.

Fritzley, V.H., & Lee, K. (2003). Do young children always say yes to yes-no questions? A metadevelopmental study of the affirmation bias. *Child Development, 74*, 1297-1313.

Fodor, J.A. (1975). *The language of thought*. New York: Crowell.

German, T.P., & Leslie, A.M. (2000). Attending to and learning about mental states. In (Eds.), P. Mitchell and K.J. Riggs, *Reasoning and the mind*. (pp. 229–252). Hove, East Sussex: Psychology Press.

German, T.P., & Leslie, A.M. (2001). Children's inferences from 'knowing' to 'pretending' and 'believing'. *British Journal of Developmental Psychology*, *19*, 59–83.

German, T.P. & Nichols, S. (2003). Children's inferences about long and short causal chains. *Developmental Science, 6*, 514-523.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*, 1–14.

Gopnik, A. (1996). The scientist as child. *Philosophy of Science*, *63*, 485-514.

Gopnik, A. & Astington, J. (1988). Children's understanding of representational change and its relation to the understanding of false belief and appearance-reality distinction. *Child Development*, *59,* 26-37.

Gopnik, A. & Wellman, H. (1992). Why the child's theory of mind *really* is a theory. *Mind and Language*, *7*, 145-171.

Gopnik, A., & Wellman, H. (1994). The theory theory. In L.A. Hirschfeld & S.A. Gelman, (Eds), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257-293). New York: Cambridge University Press.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics, Vol. 3, Speech acts* (pp. 41-58). New York: Academic.

Hale, C.M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind" a training study. *Developmental Science*, *6*, 346–359.

Happé, F. & Loth, E. (2002). 'Theory of mind' and tracking speakers' intentions. *Mind and Language, 17,* 24-36.

Joseph, R.M. (1998). Intention and knowledge in preschoolers' conception of pretend. *Child Development, 69*, 966-980.

Leslie, A.M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*, 412–426.

Leslie, A.M. (1994a). *Pretending* and *believing*: Issues in the theory of ToMM. *Cognition*, *50*, 211–238.

Leslie, A.M. (1994b). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld and S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*, (pp. 119–148). New York: Cambridge University Press.

Leslie, A.M. (2000). 'Theory of mind' as a mechanism of selective attention. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences,* 2nd Edition, (pp. 1235–1247). Cambridge, MA: MIT Press.

Leslie, A.M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, *1*, 247–254.

Leslie, A.M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, *43*, 225–251.

Leslie, A. M., German, T. P., Pollizi, P. (in press). Belief-desire reasoning as a process of selection. *Cognitive Psychology*.

Lewis, C. , & Osborne, A. (1990). Three-year-olds' problems with false belief: Conceptual deficit or linguistic artifact? *Child Development, 61*, 1514-1519.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*, 1130-1144.

Lohmann, H., Tomasello, M., & Meyer, S. (in press). Linguistic communication and social understanding. In J. Astington & J. Baird (Eds.), *Why language matters for theory of mind.* New York: Oxford University Press.

Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition, 39*, 107-127.

Moses, L.J. (2001). Executive accounts of theory of mind development. *Child Development. 72,* 688-690.

Nelson, K., Skwerer, D.S., Goldman, S., Heneseler, S., Presler, N., Walkenfeld, F.F. (2003). Entering a community of minds: An experiential approach to 'theory of mind'. *Human Development, 46*, 24-46.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J. (1995). The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition*, *57*, 241–269.

Perner J. & Lang, B. (2000). Theory of mind and executive function: Is there a developmental relationship? In S. Baron-Cohen, H. Tager-Flusberg , & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 155-186). Oxford: Oxford University Press.

Pinker, S. (1997). *How The Mind Works*. New York: Norton

Roth, D., & Leslie, A.M. (1991). The recognition of attitude conveyed by utterance: A study of preschool and autistic children. *British Journal of Developmental Psychology, 9,* 315-330.

Roth, D., & Leslie, A.M. (1998). Solving belief problems: Toward a task analysis. *Cognition*, *66*, 1–31.

Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development, 18,* 139-158.

Scholl, B.J., & Leslie, A.M. (1999). Modularity, development and 'theory of mind'. *Mind and Language*, *14*, 131–153.

Scholl, B.J., & Leslie, A.M. (2001). Minds, modules, and meta-analysis. *Child Development*, *72*, 696–701.

Siegal., M. (1997). *Knowing children: Experiments in conversation and cognition* (2nd ed.). Hove, UK: Psychology Press.

Siegal, M. (1999). Language and thought: The fundamental significance of conversational awareness for cognitive development. *Developmental Science, 2*, 1-14.

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, *38*, 1–12.

Siegal, M., Varley, R., & Want, S. C. (2001). Mind over grammar: Reasoning in aphasia and development. *Trends in Cognitive Sciences, 5,* 296-301.

Surian, L., & Leslie, A.M. (1999). Competence and performance in false belief understanding: A comparison of autistic and three-year-old children. *British Journal of Developmental Psychology*, *17*, 141–155.

Vygotsky. L.S. (1962). *Thought and Language.* Oxford, UK: Wiley.

Wellman, H.M. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

Wellman, H.M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami, (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 167-187). Oxford, UK: Blackwell.

Wellman, H.M. & Bartsch, K., (1988). Young children's reasoning about beliefs. *Cognition, 30*, 239-277.

Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory mind development: The truth about false-belief. *Child Development*, 72, 655–684.

Woolfe, T., Want, S.C., & Siegal, M. (2002). Signposts to development: Theory of mind in deaf children. *Child Development, 73,* 768-778.

Zaitchik, D. (1991). Is only seeing really believing? Sources of the true belief in the false belief task. *Cognitive Development, 6*, 91-103.

Table 1. Pattern of responses on the two false belief tasks in the standard and look first questioning conditions of Experiment 1.

| | Standard | | Look first | |
|---|---|---|---|---|
| | 3-year-olds | 4-year-olds | 3-year-olds | 4-year-olds |
| Both correct | 4 (4) | 22 (6) | 17 (12) | 40 (13) |
| One correct | 5 (4) | 9 (4) | 12 (5) | 4 (2) |
| Both incorrect | 36 (12) | 20 (10) | 16 (3) | 9 (5) |

Siegal and Beattie's (1991) results are shown in parentheses.

Table 2. Responses of the children on the false belief and true belief tasks in the standard and look first questioning conditions of Experiment 2.

| | Standard | | Look First | |
|---|---|---|---|---|
| | FB | TB | FB | TB |
| Correct | 2 (5) | 13 (11) | 7 (10) | 7 (10) |
| Incorrect | 11 (7) | 0 (1) | 2 (2) | 2 (2) |

Note. Siegal and Beattie's (1991) results are shown in parentheses.

Table 3 Number of children (percentage in parentheses) passing ‘think’ and ‘look first’ versions of the false belief tasks in Experiment 3.

| | Young 3-year-olds N=22 | Young 4-year-olds N=30 |
|---|---|---|
| Think | 3 (14%) | 17 (57%) |
| Look first | 12 (55%) | 22 (73%) |

Table 4 Contingency between answers to 'think' and 'look first' versions of the false belief task in Experiment 3.

| | | Young 3 year olds Look first question | | Young 4 year olds Look first question | |
|---|---|---|---|---|---|
| | | Pass | Fail | Pass | Fail |
| Think question | Pass | 3 | 0* | 17 | 0* |
| | Fail | 10* | 9 | 5* | 8 |
| McNemar binomial | | $p < .005$, one tailed | | $p < .05$, one tailed | |

Figure 1. Hypothesized theory-theory pattern of children's performance on standard (S) and look first (LF) false belie tasks indicating an age X condition interaction effect

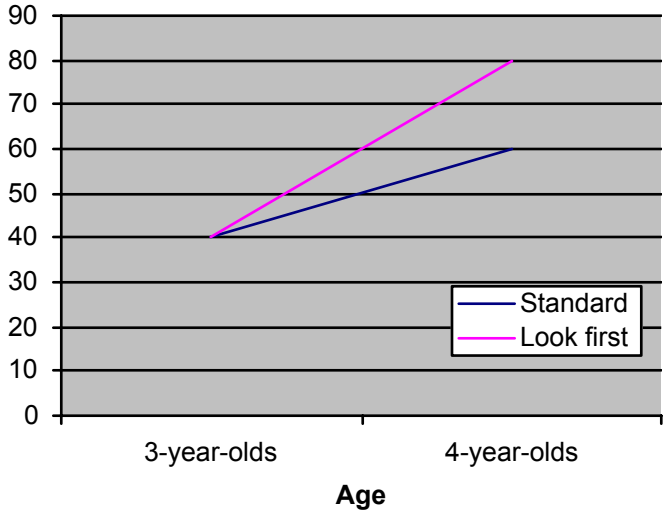



Figure 2. Hypothesized ToMM/SP pattern of childre performance on standard (S) and look first (LF) false indicating no age X condition interaction effect

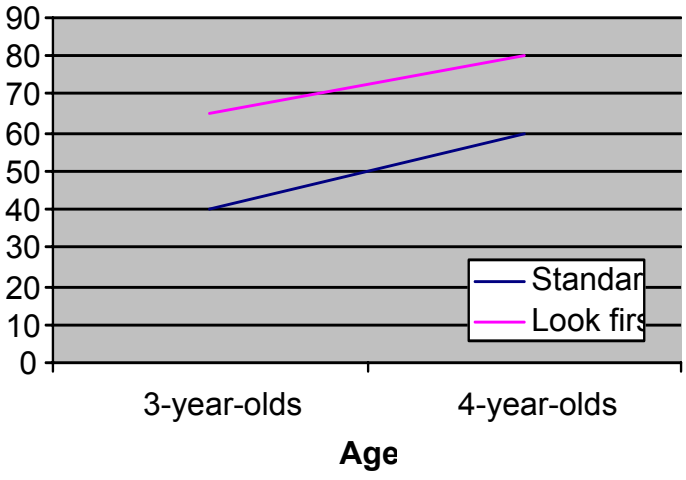


Figure 3. Children's performance on false belief tasks in the standard and look first conditions of Experiment 1

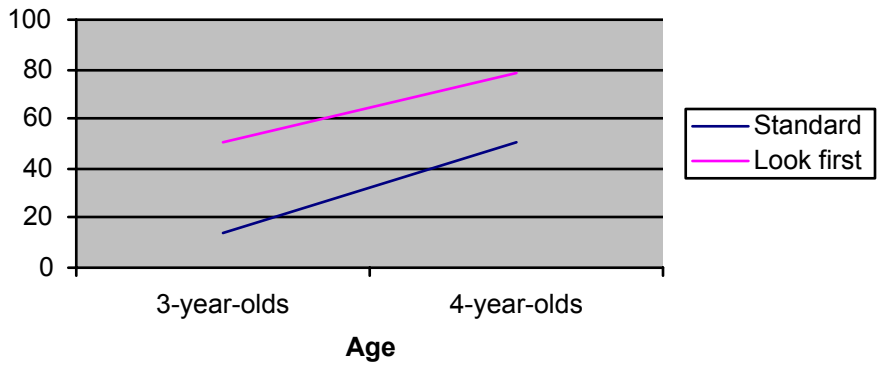


Figure 4. Performance in the standard and look first conditions of Experiment

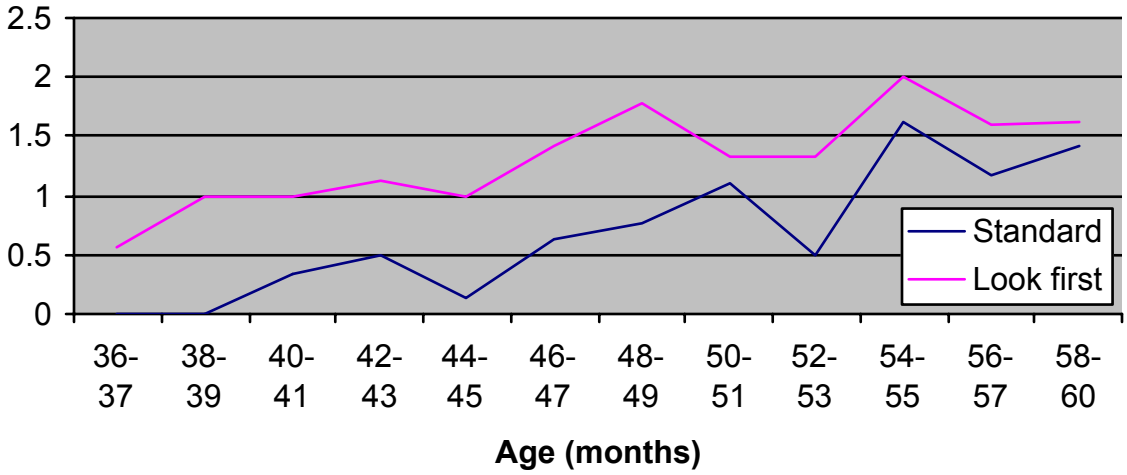


**Figure 5A: Pictures, stories and questions for true belief task in Experiment 3.**

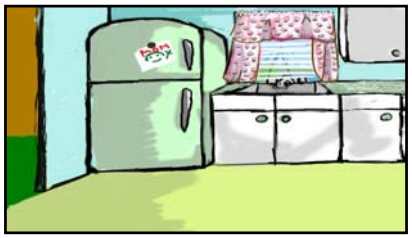
One day, Bobby is in the kitchen eating a chocolate bar.

Then his mom comes into the kitchen and says, "Bobby, put away your chocolate bar. It's time for chores."

Bobby puts his chocolate bar inside the refrigerator and…

…asks his mom what he should do next. But while he is talking to her…

…his mom takes the chocolate from the refrigerator and puts it into the candy drawer.



MQ: Where did Bobby put the chocolate in the beginning?
RQ: Where is the chocolate now?

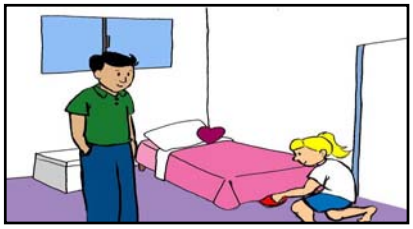
Later on, Bobby comes back into the kitchen and wants to eat his chocolate.

T: Where does Bobby think the chocolate is?
LF: Where will Bobby look first for the chocolate?

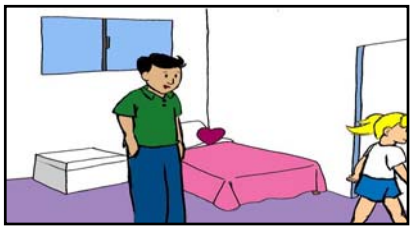
**Figure 5B: Pictures, story and questions for False belief task in Experiment 3.**

One day, Sally is playing in her room with a ball.

Then her dad comes into her room and says, "Sally, put away your ball. It's time for breakfast."

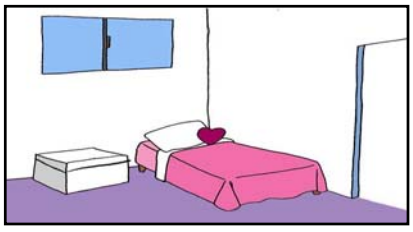


Sally puts her ball under her bed and…



…goes down for breakfast. But while she is away…

…her dad takes the ball from under the bed and puts it into the toy chest.



MQ: Where did Sally put the ball in the beginning?
RQ: Where is the ball now?

Later on, Sally comes back into her room and wants to play with her ball.
T: Where does Sally think the ball is?
LF: Where will Sally look first for her ball?